BBC RD 1994/10

BBC

Research and Development Report

DIGITAL VIDEO COMPRESSION:
Standardisation of scalable coding schemes

P.N. Tudor, M.A. and N.D. Wells, B.A., D.Phil

Research and Development Department,
Engineering Division
THE BRITISH BROADCASTING CORPORATION

BBC RD 1994/10

DIGITAL VIDEO COMPRESSION: STANDARDISATION OF SCALABLE CODING SCHEMES

P.N. Tudor, M.A. and N.D. Wells, B.A., D.Phil.

Summary

Scalable coding schemes for bit-rate reduction of digital TV, including HDTV, are being considered in order to provide for compatibility with lower resolution scanning formats and to provide for graceful degradation of the picture quality in the presence of transmission errors.

This Report describes some of the scalable coding schemes considered within the Moving Pictures Expert Group (MPEG) of the International Standards Organisation (ISO) and discusses the advantages and disadvantages of the different schemes which may be classified as frequency scalable, spatial scalable and SNR scalable. The reasons for selecting particular scalable schemes for inclusion in the MPEG-2 standard are described.

Issued under the Authority of

General Manager
Research and Development Department

Research and Development Department
Engineering Division
BRITISH BROADCASTING CORPORATION

(T-30)

1994

DIGITAL VIDEO COMPRESSION: STANDARDISATION OF SCALABLE CODING SCHEMES

P.N. Tudor, M.A. and N.D. Wells, B.A., D.Phil.

1. Background 1

2. Non-scalable Algorithm 1

3. Scalable Schemes 2

   3.1 Frequency scalable 2

   3.2 Spatial scalable 5

   3.3 Signal to noise ratio (SNR) scalable 5

4. Chrominance Scalability 7

5. Standards Conversion 8

6. Discussion 9

7. Performance 9

8. Conclusions 9

9. References 10

DIGITAL VIDEO COMPRESSION: STANDARDISATION OF SCALABLE CODING SCHEMES

P.N. Tudor, M.A. and N.D. Wells, B.A., D.Phil.

1. BACKGROUND

A scalable video coding scheme is one which produces a bitstream that can be usefully decoded in parts. Decoding part of the bitstream will produce decoded pictures with less resolution or more noise than pictures decoded using the complete bitstream.

Scalable video coding schemes have a number of applications in digital communication systems. For telecommunications, scalable coding can provide compatibility with existing video transmission standards and assist concealment of transmission errors; for computer based video systems, scalable coding can generate a low resolution video preview without decoding a full resolution picture; for broadcasting applications, scalable coding can provide a range of picture qualities suited to different types of receiver and provide for graceful degradation of the picture quality in the presence of transmission errors.

Different applications have different quality requirements for the different layers within the bitstream. For example, in a broadcasting application in which a lower layer is being received continuously by portable receivers, optimisation of the quality of the lower layer may be as important as the optimisation of the quality of the higher layer. However, in future computer applications involving the retrieval of stored video information, having simple methods for the rapid decoding and viewing of low resolution preview pictures may be more important than maintaining the quality of these pictures.

2. NON-SCALABLE ALGORITHM

The scalable coding schemes considered within MPEG[1] are all extensions of the non-scalable codec shown schematically in Fig. 1*. The coder generates a motion-compensated prediction for a picture based on previous locally-decoded pictures. The locally-decoded picture is reconstructed exactly as it would be in the decoder. The difference between the prediction and the source picture is transformed using a Discrete Cosine Transform (DCT) which exploits the correlation between elements of each picture. The resulting DCT coefficients are quantised according to visibility weightings, variable length coded and transmitted to the decoder.

Within the decoder, a motion-compensated prediction based on previously decoded pictures is formed, and is identical to the one in the coder loop. The prediction error sent from the coder is added to this prediction to produce the decoded output.

To allow the decoder to start decoding, the coder periodically transmits an 'intra' picture that does not require a motion-compensated prediction in the decoder. Once this picture is decoded, the decoder is able to generate the motion-compensated predictions necessary to decode future pictures.

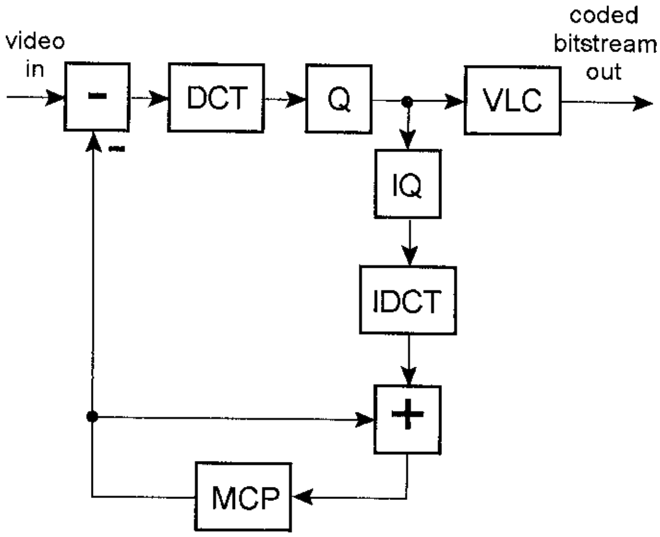


Fig. 1(a) - A non-scalable coder.

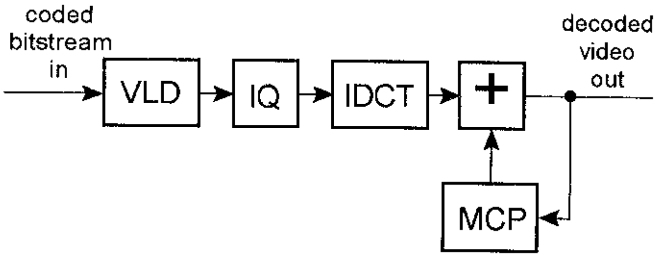


Fig. 1(b) - A non-scalable decoder.

* In the codec diagrams, the following abbreviations are used:

DCT	discrete cosine transform	Q	quantisation	VLC	variable length coder	MCP	motion compensated prediction	↓	down conversion
IDCT	inverse discrete cosine transform	IQ	inverse quantisation	VLD	variable length decoder	▟	DCT coefficient splitting or summing	↑	up conversion

3. SCALABLE SCHEMES

The scalable coding schemes considered within MPEG extend the basic codec architecture to provide layers of different resolution and levels of different picture quality (within a layer). The schemes can be classified as frequency scalable, spatial scalable and noise (SNR) scalable.

3.1 Frequency scalable

Frequency scalable schemes are characterised by the use of the DCT to split the signal into different resolution layers and by the prediction of DCT coefficients between layers. The principle underlying this approach is outlined in Fig. 2 which may be explained as follows.

Each block of the image, consisting of 8 pixels by 8 lines, is transformed using the DCT. This process gives 64 coefficients which may be transformed back to the original image using the inverse DCT process. The DCT is closely related to the Fourier Transform and we find, for example, that high horizontal frequencies present in the block tend to give high horizontal frequency DCT coefficients. Similarly, high vertical frequencies in the image block tend to give high vertical frequency coefficients.

Having transformed a block, if all the coefficients are replaced by zero except the 4 × 4 lowest frequency coefficients, high frequency information is removed (in the main) from the block. Then, if a full 8 × 8 inverse transform is performed (including the 48 zero coefficients), an approximation to a low pass filtered version of the original image block is obtained. However, it is also possible to do a 4 × 4 inverse transform using only the 4 × 4 non-zero coefficients and generate a 4 × 4 pixel image block. This 4 × 4 image block corresponds (approximately) to a low-pass filtered and subsampled version of the original 8 × 8 image block. Fig. 3 shows a two-layer frequency scalable video coder and decoder which uses this principle.

The coder loop operates in a manner similar to the non-scalable codec, using an 8 × 8 DCT to split the prediction error signal into 64 coefficients. A block of 4 × 4 low frequency coefficients is removed from the DCT coefficient matrix, quantised, and transmitted in a lower layer. The upper layer transmits refinements to these coefficients and the remaining high frequency coefficients. By adjusting the quantisation in each layer, the distortion and bit rate of the layers can be balanced by the coder. The best picture quality in the upper layer is achieved when minimal refinement of the low frequency coefficients is required. In this case, the upper layer picture quality is similar to the non-scalable case. For this to be achieved, a high proportion of the bit rate must be used for the lower

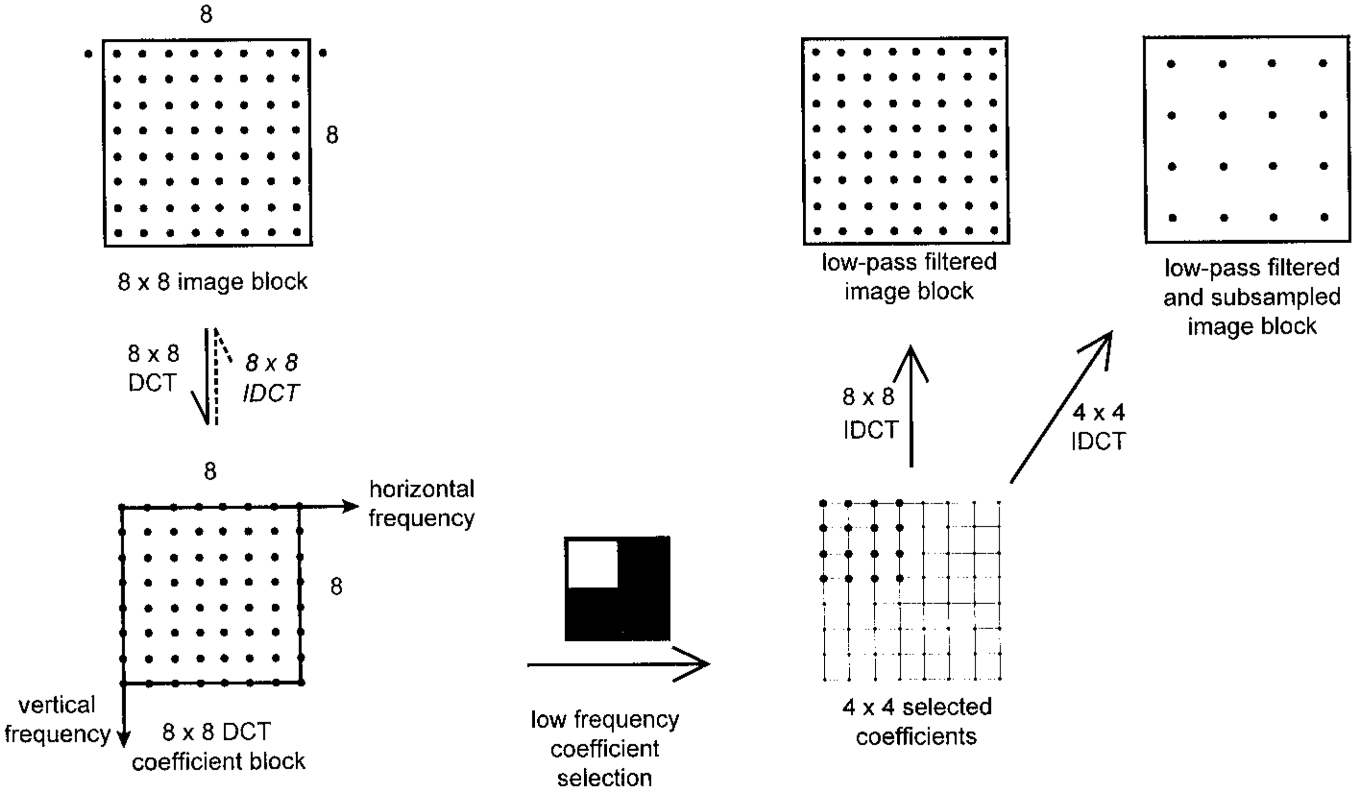

8
8
8 x 8 image block
8 x 8 DCT
8 x 8 IDCT
8
horizontal frequency
8
vertical frequency
8 x 8 DCT coefficient block
low frequency coefficient selection
4 x 4 selected coefficients
8 x 8 IDCT
4 x 4 IDCT
low-pass filtered image block
low-pass filtered and subsampled image block

Fig. 2 - Principles of frequency scalable coding.

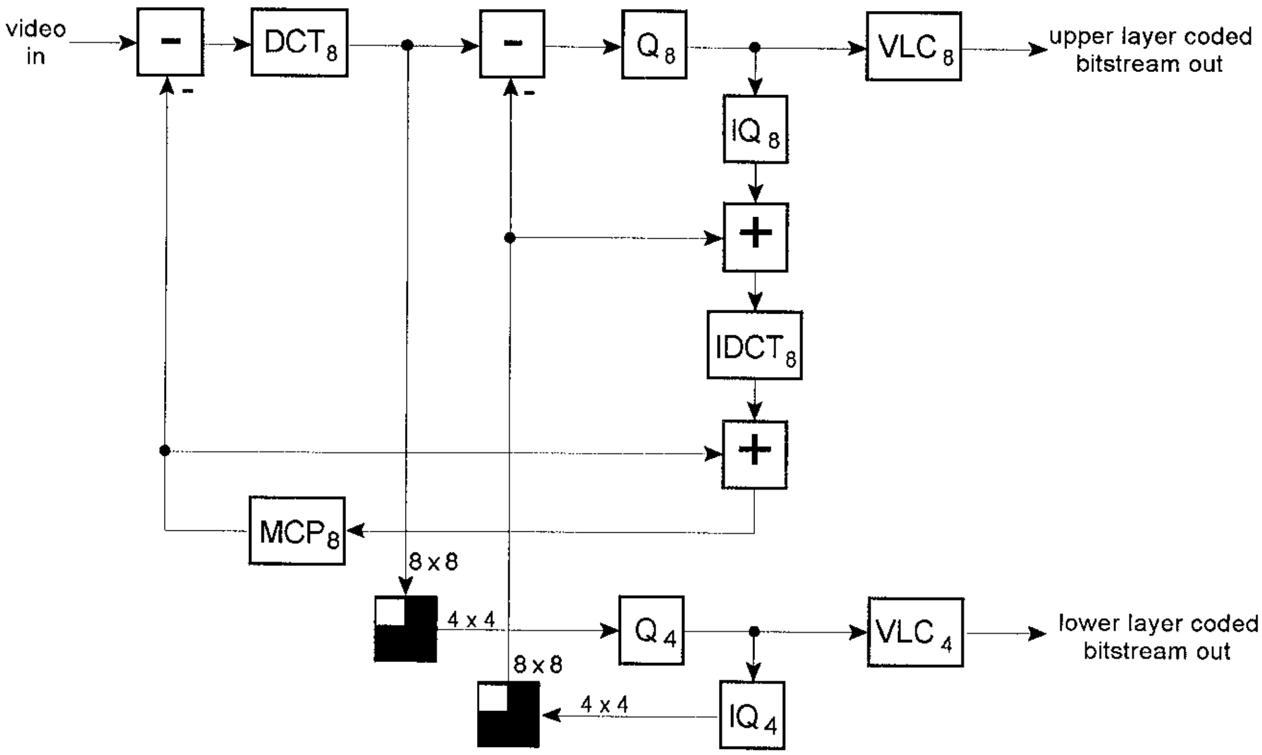


Fig. 3(a) - A two-layer frequency-scalable coder.

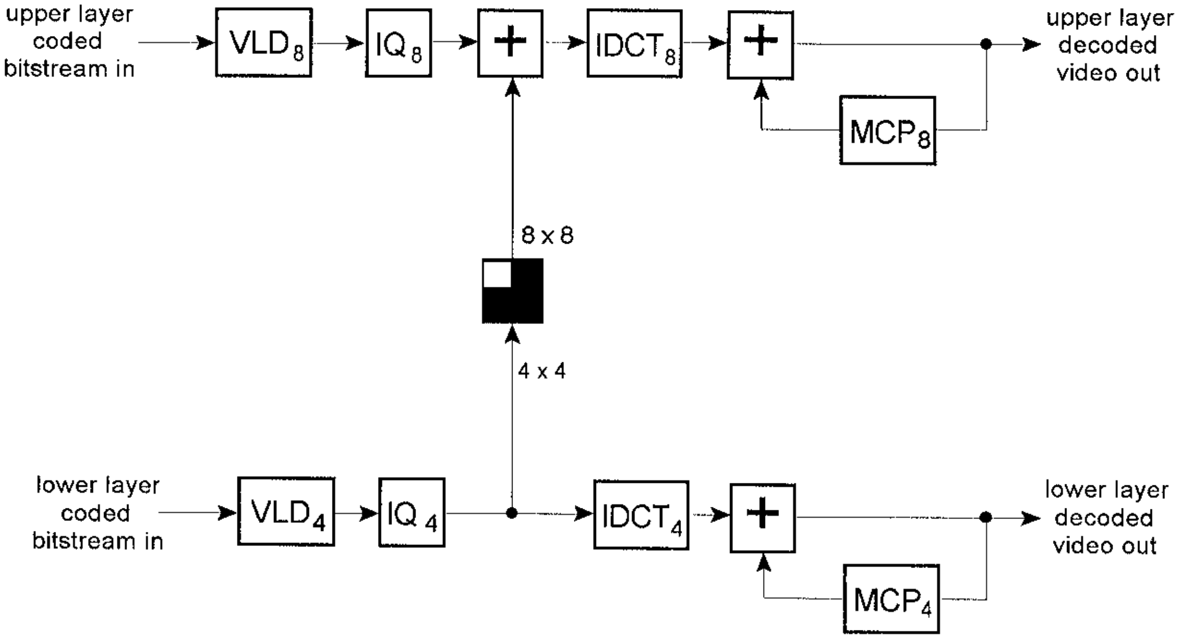

upper layer coded bitstream in
VLD8
IQ8
+
IDCT8
+
upper layer decoded video out
MCP8
8 x 8
4 x 4
lower layer coded bitstream in
VLD4
IQ4
IDCT4
+
lower layer decoded video out
MCP4

Fig. 3(b) - A two-layer frequency-scalable decoder.

layer. If the bit rate in the lower layer is reduced, additional bit rate is required in the upper layer for the refinement of the low frequency coefficients and the quality of the upper layer picture tends to be reduced.

In an upper layer decoder, the quantisation refinement is added to the 16 coefficients of the lower layer and these are merged with the other 48 higher frequency coefficients. Then, the full 8 × 8 inverse DCT is performed to give a full resolution prediction error, which is added to the full resolution motion compensated prediction to give the upper layer output. In a lower layer decoder, only the lower layer coefficients are received and a 4 × 4 inverse DCT is used to reconstruct a low resolution picture. The decoded lower layer has half the resolution and sample rate of the upper layer in both horizontal and vertical directions.

An advantage of this frequency-scalable scheme is that only one motion compensated prediction loop is needed to decode each layer. However, there are a few limitations to such a scalable scheme. Firstly, the down-conversion to the lower layer is by low frequency coefficient selection which is far from optimum when the upper and lower layers are interlaced, as is the case for HDTV in the upper layer and TV in the lower layer. This is because the high frequency coefficients are actually carrying information about the motion between adjacent interlaced fields. Methods to optimise the lower-layer quality have been tried, such as selecting coefficients from field blocks rather than frame blocks and such as retaining some high vertical frequency coefficients in the 4 × 4 coefficient selection. However, such methods have not managed to remove visible down-conversion artifacts from the lower-resolution picture.

Secondly, the quality of the decoded lower layer is degraded by what is termed 'drift' in the lower layer. Drift is a progressive deterioration in the quality of the decoded picture which arises when the prediction in the decoder is not being constructed in exactly the same way as the prediction in the coder. In this case, the lower-layer decoder loop is operating at a different resolution from the coder loop and these prediction loops are not exactly equivalent for the low frequency information. The predictions differ because the 8 × 8 and 4 × 4 DCTs do not split the signal perfectly into different frequency bands and because the motion vectors may not be applied to the same accuracy in the upper- and lower-resolution decoder loops. The seriousness and visibility of drift depends on many factors, including the scene content, the time interval between intra-coded pictures, the display size and the application for the decoded pictures.

The quality of the down-conversion can be improved and drift problems in the lower layer avoided by having a dual-loop coder as shown in Fig. 4. In this case, the down-conversion is performed

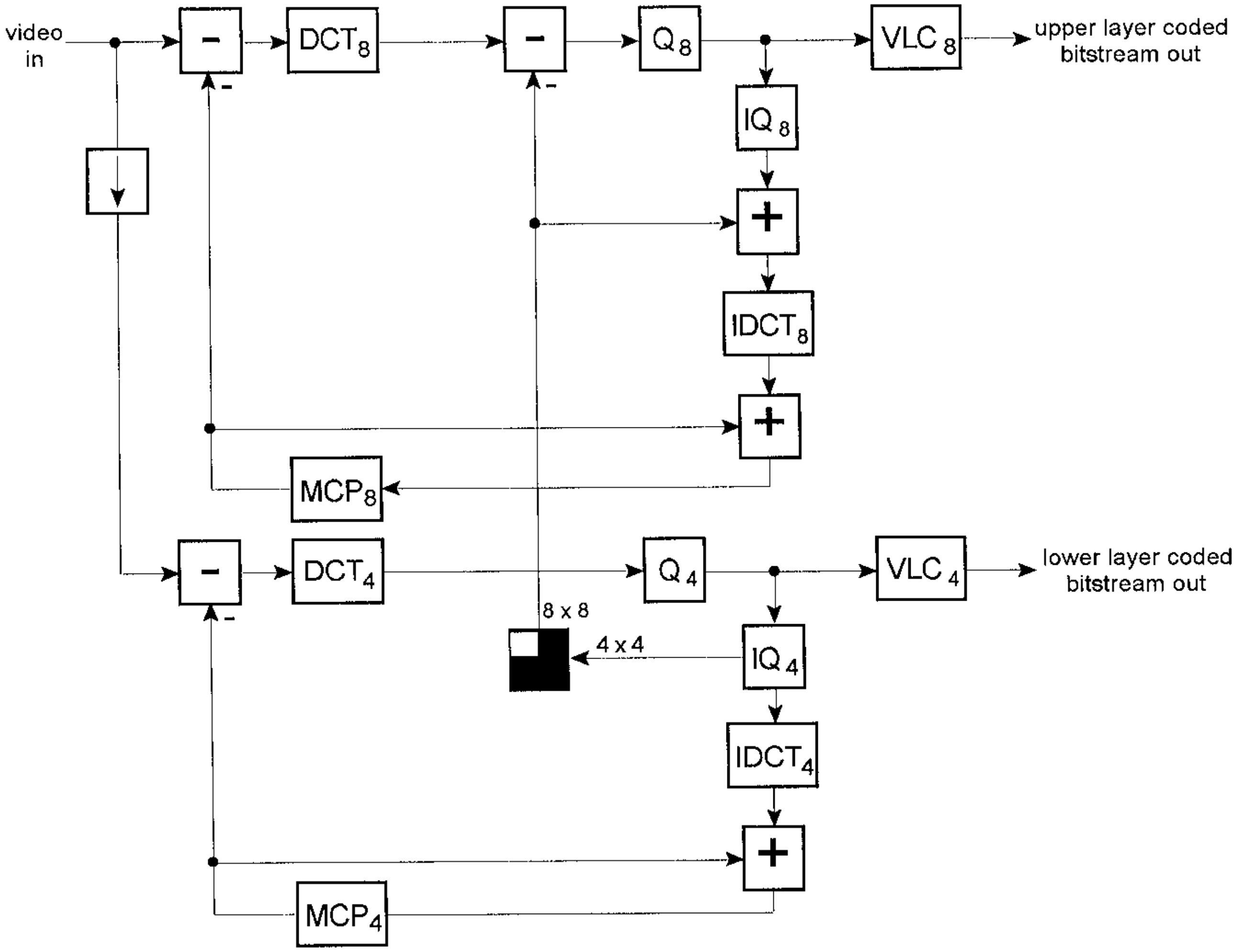


Fig. 4 - Two-loop frequency-scalable coder with external down-conversion.

externally and can be optimised independently of the coding process. The lower-layer loop then uses a 4 × 4 DCT process which matches that used in the lower-layer decoder and so there will be no drift in the lower layer decoder. The 16 coefficients per block from the lower layer are used to predict the 16 low frequency coefficients in the upper layer. The decoder is the same as the one shown in Fig. 3(b).

The disadvantages of using a dual-loop coder are that the complexity of the coder is increased and the upper-layer decoded picture quality is reduced. This latter effect is because the lower-layer loop coefficients make a less efficient prediction for the lower frequency coefficients of the upper-layer loop, than in the single loop frequency-scalable case.

In most proposed frequency-scalable schemes, the spatial resolution changes by a factor of two between layers. Changes in resolution by factors other than two are not normally considered. Note that in frequency-scalable schemes, lower-layer decoders would be required to support 4 × 4 inverse DCT (and 2 × 2 inverse DCT for three resolution layers), rather than the 8 × 8 inverse DCT used in non-scalable MPEG decoders.

3.2 Spatial scalable

Spatial-scalable schemes are characterised by the use of up-converted pictures from a lower layer as a prediction in a higher layer. Fig. 5 *(overleaf)* shows a block diagram of a two-layer spatial-scalable codec. There is no restriction on the ratio of spatial or temporal sampling frequencies in the different layers, except for the complexity of the down- and up-conversion processes.

In the scheme shown in Fig. 5, the prediction in the upper layer can be taken as the motion-compensated prediction from the upper-layer loop (termed the *temporal* prediction) or the up-converted decoded picture from the lower-layer loop (termed the *spatial* prediction) or a weighted combination of these two predictions. This selection is performed by the adaptive weighting function, W, in Fig. 5.

In such a scheme, there is no drift in the decoded lower layer and the down-conversion to the lower layer can be optimised as required. Also, only 8 × 8 inverse DCT is required in the lower layer decoder. The quality of the lower layer is not reduced in any way by the scalability, but the quality of the upper layer is reduced because the bit rate used on the lower layer is not completely compensated for by the improved efficiency of prediction available in the upper layer.

3.3 Signal to noise ratio (SNR) scalable

SNR-scalable schemes quantise the coefficients to different levels of accuracy in the different SNR levels. Fig. 6 shows a block diagram of a two-level SNR-scalable video codec.

The codec operates in a similar manner to the single-level codec, with the addition of an extra quantisation stage. The lower-level coder quantises the DCT coefficients to a given accuracy and transmits these quantised coefficients in the lower-level bitstream. The quantisation error introduced by the lower-level quantiser is then requantised and sent in the upper level as a refinement to the lower-level coefficients.

For an upper-level decoder, both levels must be received and the upper-level coefficient refinements are added to the lower-level coefficient values following inverse quantisation. The summed coefficients are then decoded using a single-loop decoder. The lower-level decoder receives just the coefficient values in the low-level bitstream and decodes these directly using a single-loop decoder. In the upper-level bitstream, as little data as possible is transmitted, apart from the coefficient enhancement values. For example, motion vector information is only sent in the lower-level bitstream.

Note that the lower-level decoder output will have some drift because the prediction loop has at its input coarsely-quantised coefficients, whereas the coder loop (and the upper-level decoder loop) use the more finely-quantised coefficients. However, experiments have shown that this drift in the lower SNR level is not very visible and the resulting quality is acceptable for many applications. Drift in the lower level can be corrected by using a dual-loop coder, in a manner similar to that shown in Fig. 4 for a frequency scalable coder, except that lower- and upper-level loops work at the same picture resolution. However, a dual-loop coder is not as efficient as a single-loop coder in terms of the upper-level picture quality at a given bit rate.

Fig. 7 *(page 8)* shows an alternative SNR scheme, in which the lower-level decoder drift has been corrected by using the lower-level locally decoded picture in the coder loop. Thus, the lower-level decoder now derives its prediction in the same manner as the coder loop. The upper-level decoder reconstructs the lower-level picture and adds to this a correction derived from the coefficient refinement sent in the upper level, as shown in Fig. 7(b). For the SNR scheme of Fig. 7, the quality of the upper level is reduced compared to the SNR scheme in Fig. 6, because the quality of the prediction is based on the

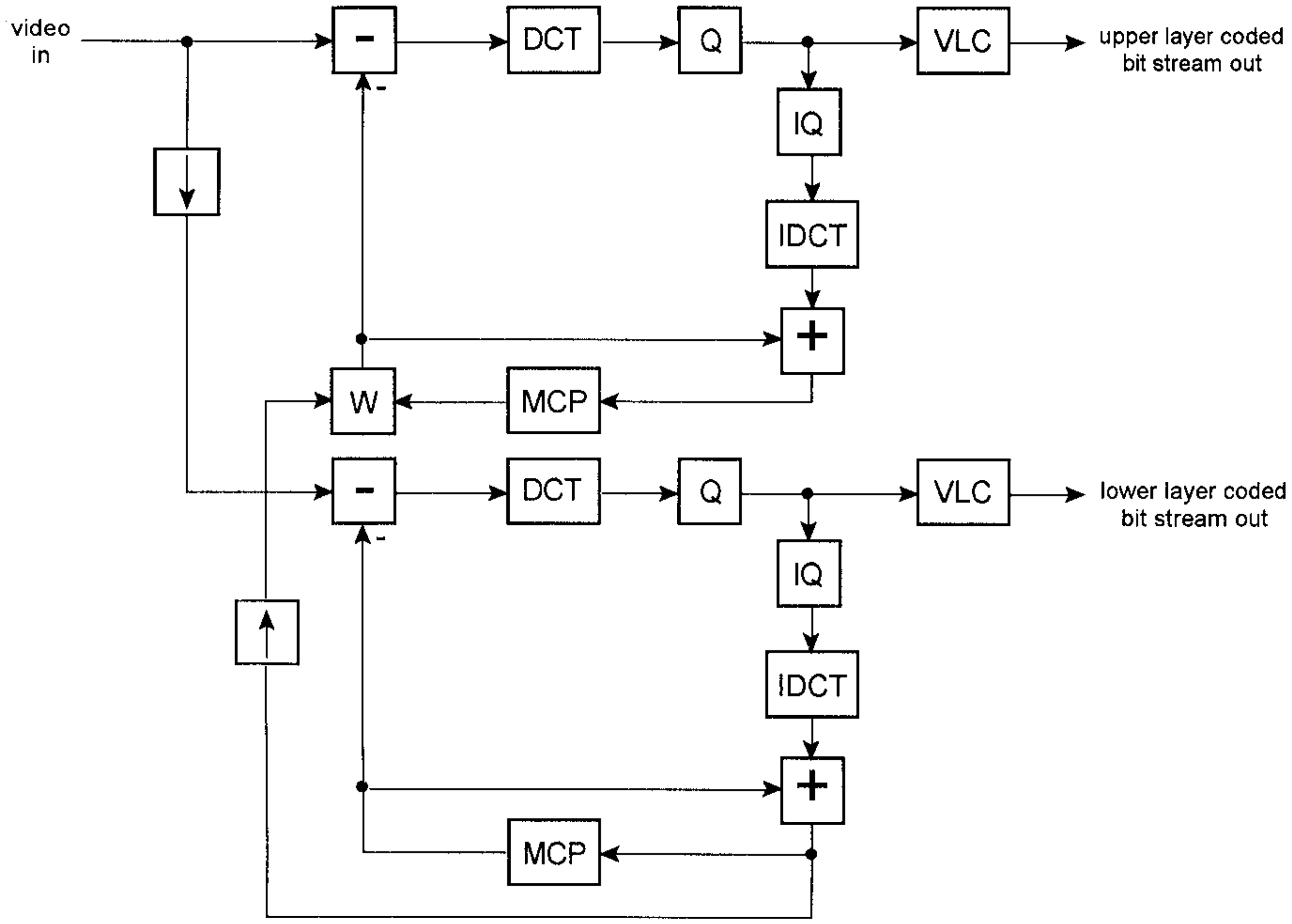

video in
-
DCT
Q
VLC
upper layer coded bit stream out
IQ
IDCT
+
W
MCP
-
DCT
Q
VLC
lower layer coded bit stream out
IQ
IDCT
+
MCP

Fig. 5(a) - A two-layer spatial-scalable video coder.

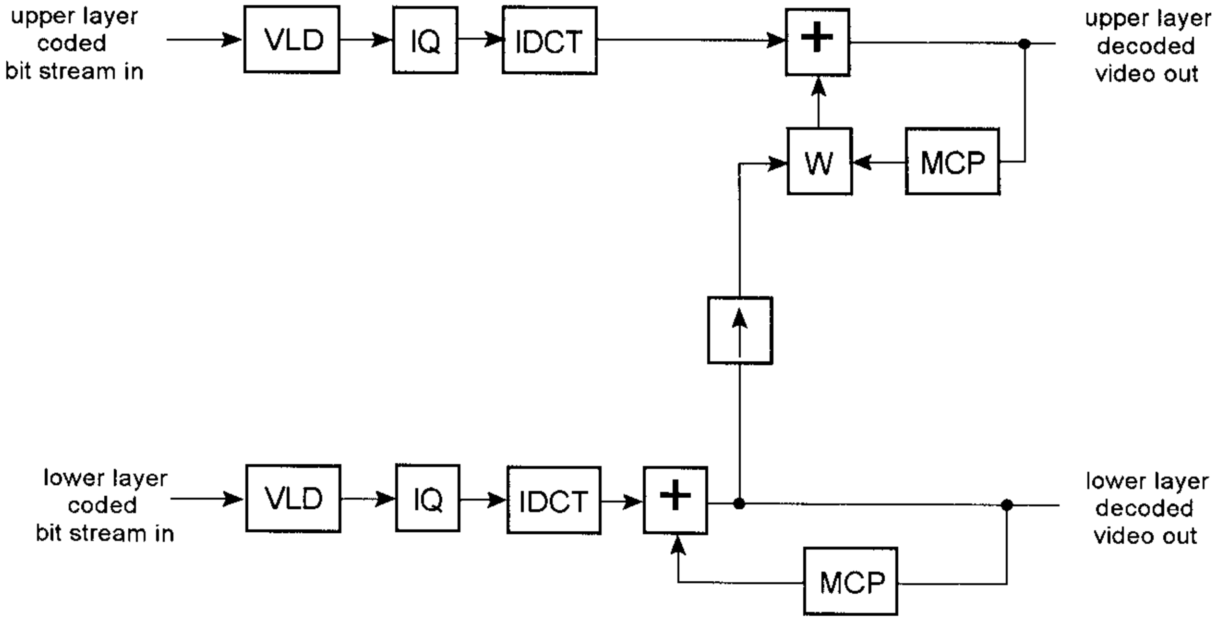
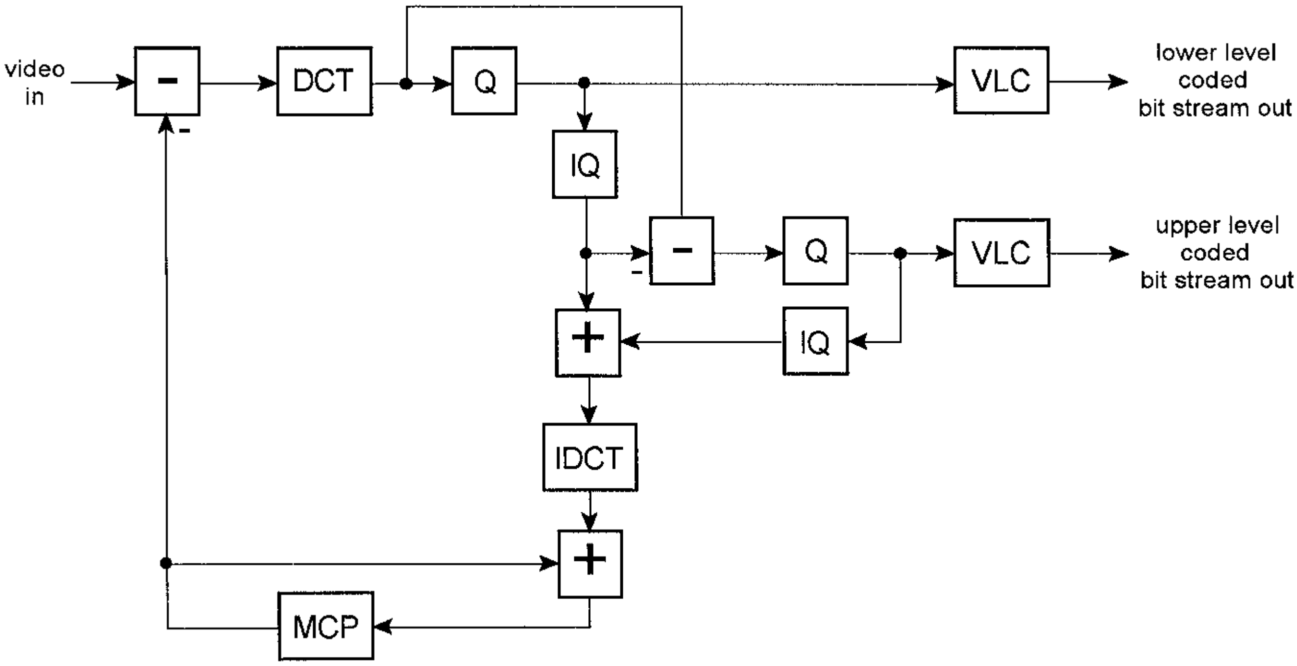
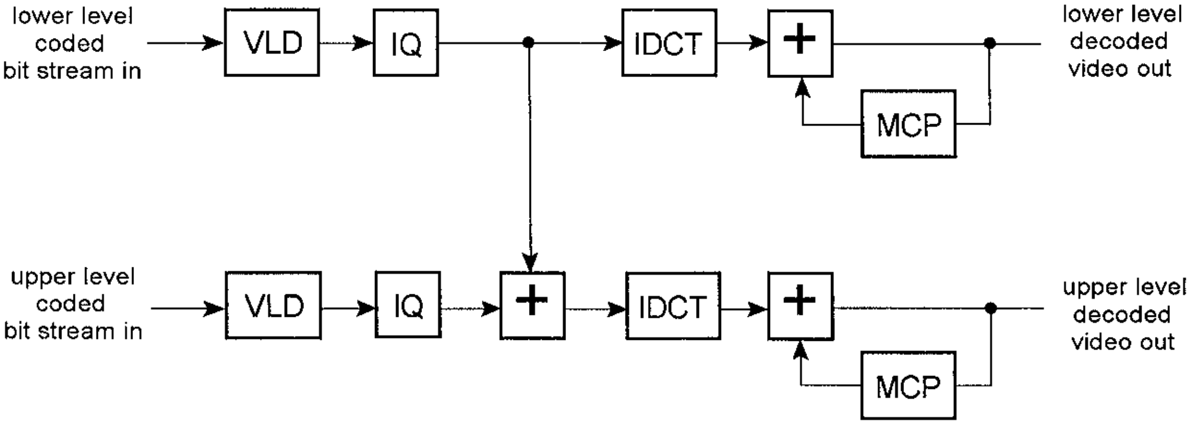

upper layer coded bit stream in
VLD
IQ
IDCT
+
upper layer decoded video out
W
MCP
lower layer coded bit stream in
VLD
IQ
IDCT
+
lower layer decoded video out
MCP

Fig. 5(b) - A two-layer spatial-scalable video decoder.



video in
-
DCT
Q
VLC
lower level coded bit stream out
IQ
-
-
Q
VLC
upper level coded bit stream out
+
IQ
IDCT
+
MCP

Fig. 6(a) - A two-level SNR-scalable video coder.



lower level coded bit stream in
VLD
IQ
IDCT
+
lower level decoded video out
MCP
upper level coded bit stream in
VLD
IQ
+
IDCT
+
upper level decoded video out
MCP

Fig. 6(b) - A two-level SNR-scalable video decoder.

lower quality of the lower-level signal rather than the upper-level signal. Consequently, this latter SNR scheme was not chosen for standardisation by MPEG.

4. CHROMINANCE SCALABILITY

MPEG is directed, in the main, towards systems in which the vertical sampling rate and resolution for the chrominance is half that of the luminance. However, for some studio-based applications it may be desirable to maintain the full chrominance resolution of the source pictures.

Spatial scalability was investigated as a means of sending an enhancement signal carrying the additional chrominance resolution. It was found that there was little to be gained by going to the complexity of spatial-scalable coding compared with simulcasting the whole of the full-resolution chrominance component in addition to the lower-resolution chrominance component. However, it was found efficient to predict the DC coefficients of the high-resolution chrominance blocks using the appropriate DC coefficients of the lower-resolution chrominance blocks. As little additional information as possible is sent in the enhancement layer and the syntax used to describe SNR scalability is used to transmit the chrominance 'simulcast' enhancement.

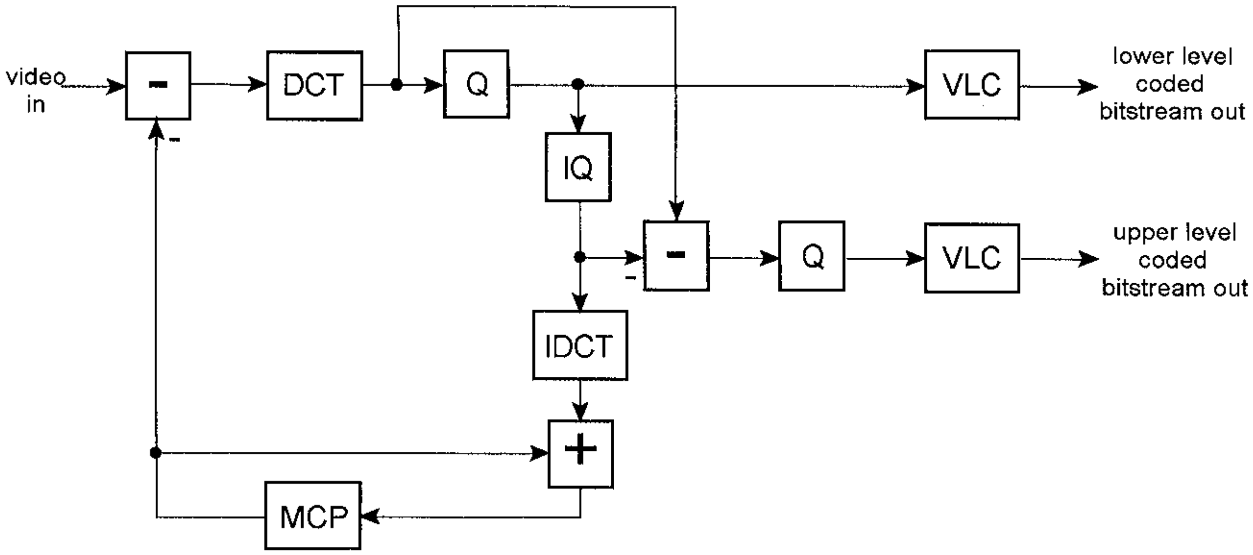

video in
-
DCT
Q
IQ
VLC
lower level coded bitstream out
-
Q
VLC
upper level coded bitstream out
IDCT
+
MCP

Fig. 7(a) - Two-level SNR-scalable coder, without drift in the lower level.

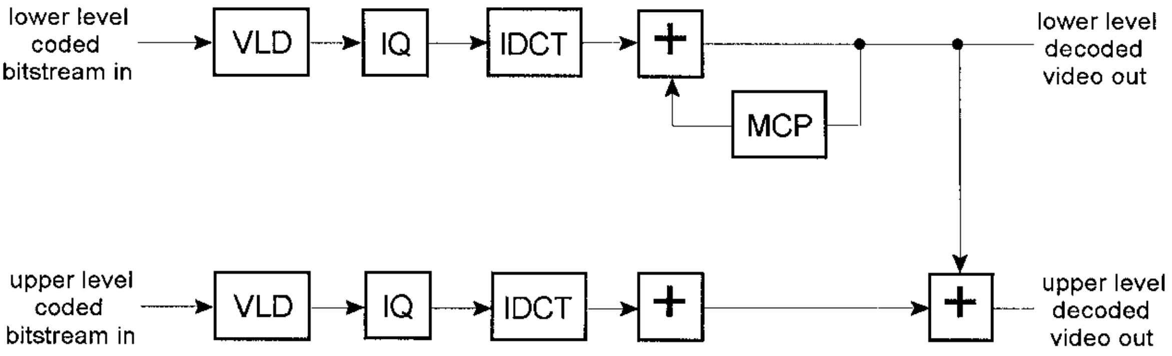

lower level coded bitstream in
VLD
IQ
IDCT
+
MCP
lower level decoded video out
upper level coded bitstream in
VLD
IQ
IDCT
+
+
upper level decoded video out

Fig. 7(b) - Two-level SNR-scalable decoder, without drift in the lower level.

5. STANDARDS CONVERSION

Satisfactory performance of resolution-scalable schemes depends on good quality standards conversion. Conversion between existing broadcast standards is problematical because of the interlaced scanning. Optimum performance for stationary pictures can be obtained if information from both input fields is used to derive each output field. However, with fast movement, better results are obtained if each output field is derived using information from only one input field because of the large changes in picture detail between fields.

Good quality conversion can be obtained using systems which adapt between inter-field and intra-field interpolation according to the amount of movement in the sequence. However, such methods can introduce visible switching effects.

More consistent results can be obtained using a non-adaptive system based on a combination of vertical and temporal interpolation[2]. The interpolation process can be described as a vertical-temporal filter removing unwanted spectral components from the input signal which would give rise to undesirable alias components at the output scanning standard. An important feature of the vertical-temporal interpolator is that the sum of the components from non-central fields is equal to zero; this ensures good conversion of moving objects.

Three-field vertical temporal filters can give high-quality conversion between interlaced standards. The down-conversion process does not need to be defined for scalable schemes within the MPEG standard. However, in the case of spatial scalability, where an up-converter is required in each coder and decoder, the up-conversion process must be precisely defined. The choice of up-converter was chosen while taking the following factors into account:

i) The up-converter should be as simple as possible for mass-produced decoders.

ii) The up-converter should be intra-field or intra-frame in order to avoid complications introduced by possible frame reordering for transmission. The frame order for transmission is changed when bi-directional prediction is used.

iii) The up-conversion process must be capable of simple extension to cases where the upsampling ratio is other than 2:1.

iv) The upper-layer picture quality is not critically dependent on the quality of the up-conversion.

v) The up-conversion should provide as good an up-converted picture quality as possible because, when graceful degradation is required in the presence of transmission errors, the up-converted lower layer is used as a replacement for the decoded upper-layer picture.

The up-converter defined for spatial scalability within MPEG combines intra-frame up-conversion with linear interpolation which can be applied for an arbitrary up-conversion ratio.

6. DISCUSSION

Frequency- and spatial-scalable schemes provide layers of different resolution, and SNR-scalable schemes provide levels of different picture quality within a layer.

For broadcasting applications, the picture quality of the lower layer is important and, for interlaced picture formats, external down-conversion is required which filters the picture in an optimum way according to the scan conversion. Frequency- and spatial-scalable schemes support external down-conversion, and algorithmically become similar with its inclusion. Also, the performance of the two schemes then becomes very similar in terms of the picture quality in both layers.

It is expected that by the time an HDTV service starts there will be a large number of MPEG-2 main profile TV decoders installed and compatibility with these will be important. MPEG main profile TV decoders support only 8 × 8 inverse DCT. Spatial scalability is then the only option for providing resolution scalability because a frequency-scalable scheme would require 4 × 4 or 2 × 2 inverse DCT processing which would not be compatible with main profile decoders. Therefore, the spatial-scalable scheme was the one chosen for MPEG-2 to provide for resolution scalability.

For computer-based video systems that require resolution scalability (in order to offer rapid decoding of picture previews), the quality of the lower layer may be less important. In such cases, frequency-scalable schemes can offer a low-complexity solution whilst maintaining high quality in the upper layer. If required, the SNR tools can be used to provide frequency-scalable functionality.

7. PERFORMANCE

One method of offering a range of different picture resolutions and qualities is to use the 'simulcast' method. With this method, independent coders operate in parallel to produce independent bitstreams intended for independent decoders. Each decoder uses only one bitstream and no attempt is made to use information intended for one decoder in another. Simulcast represents the lower coding performance threshold against which scalable schemes can be compared.

Using a two-layer spatial-scalable scheme with interlaced HDTV and TV, at bit rates of 14 Mbit/s and 6 Mbit/s respectively, the HD layer SNR is improved by between 0.5 dB and 1.5 dB relative to simulcast. The higher figure is approximately half way between the quality obtained by simulcasting and the non-scalable case where the entire bit rate is used for the HD layer. The improvement due to spatial scalability increases in scenes containing complex motion, for which the upper-layer temporal prediction is poor. The main benefit of spatial scalability is in intra-coded pictures where no temporal prediction is available. If a significant proportion of the available bit rate has been spent on coding the lower layer, then it is worthwhile to use spatial scalable coding to improve the efficiency of the upper-layer coding.

SNR scalability works well. In a two-level SNR-scalable scheme, the loss of picture quality in the upper level is only slight compared to a non-scalable scheme using the entire bit rate. In the lower level, drift is barely perceptible and the picture quality is similar to that of a non-scalable scheme operating at the lower level bit rate.

8. CONCLUSIONS

The scalable coding schemes considered by MPEG have been discussed. Frequency- and spatial-scalable schemes offer scalable resolution, and SNR scalability offers scalable picture coding quality.

For broadcasting applications, resolution scalability is best achieved using the spatial-scalability method. This permits a high-quality TV layer, supports a wide distribution of bit rates between the layers, and is compatible with non-scalable main profile MPEG decoders in the lower layer.

9. REFERENCES

1. ISO/IEC, 1994. Generic Coding of Moving Pictures and Associated Audio, DIS 13818-2 video. JTC/SC29/WG11, Paris, March 1994.

2. DEVEREUX, V.G., 1992. Standards conversion between 1250/50 and 625/50 TV systems. IBC 1992. IEE Conference Publication No. 358, pp. 51-53.

Published by BBC RESEARCH & DEVELOPMENT DEPARTMENT, Kingswood Warren, Tadworth, Surrey, KT20 6NP